



Applying new models of data stewardship to health and care data



This is licensed under a Creative Commons Attribution-ShareAlike 4.0 International license

Open Data Institute

Contents

Contents	1
About	3
Foreword	3
Executive summary	4
Recommendations	5
Introduction	7
Background	9
What do we mean by evaluation?	9
What do we mean by new models of data stewardship?	10
Three use cases	12
Relevance for Covid-19	13
Use case #1: Digital-first primary care	15
Use case description and background	15
Findings	15
What an evaluation would aim to assess or demonstrate	15
Data an evaluation would need	16
Challenges or barriers to accessing necessary data	16
Recommendations and next steps	16
Embed mandatory access to data for evaluation purposes in procurement contracts	17
Enable institutional efforts to improve access to NHS data to support technology evaluation	17
Engage with existing initiatives and consider piloting a data institution to facilitate access to NHS data for the purpose of evaluations	18
Explore the potential for existing organisations to take on new data stewardship roles	19
Advocate for technologies to be implemented in a more evidence-based way, with early input from evaluators	19
Covid-19 context	20
Use case #2: Online misinformation and vaccine hesitancy	21
Use case description and background	21
Findings	21
What an evaluation would aim to assess or demonstrate	21
Data an evaluation would need	22
Challenges or barriers to accessing necessary data	22
Recommendations and next steps	23
Explore the use of data portability to collect contextual data from offline sources	23
Explore the use of plugins and data portability to collect data from online sources	24

Consider piloting a data institution aimed at facilitating access to data held by online platforms	25
Engage with stakeholders to identify their views and needs	26
Rethink the scope of evaluations on this topic	26
Covid-19 context	27
Use case #3: Patient flow automation	28
Use case description and background	28
Findings	29
What an evaluation would aim to assess or demonstrate	29
Data an evaluation would need	29
Challenges or barriers to accessing necessary data	29
Recommendations and next steps	30
Consider piloting a data institution aimed at enabling comparative and long-term evaluations	30
Investigate the form and scale of a data institution	31
Covid-19 context	32
Conclusions and recommendations	33
Recommendations for evaluators	33
Recommendations for funders	34
Recommendations for innovators	34
Recommendations for health and care providers	35
Recommendations for patient and practitioner groups	35
Appendices	36
Appendix 1: Project methodology	36
Appendix 2: Detailed findings	37
Use case #1: Detailed findings	37
What an evaluation would aim to assess or demonstrate	37
Data an evaluation would need	37
Challenges or barriers to accessing necessary data	38
Use case #2: Detailed findings	39
What an evaluation would aim to assess or demonstrate	39
Data an evaluation would need	40
Challenges or barriers to accessing necessary data	40
Use case #3: Detailed findings	41
What an evaluation would aim to assess or demonstrate	41
Data an evaluation would need	42
Challenges or barriers to accessing necessary data	42

About

This report has been researched and produced by the Open Data Institute and published in July 2020. It was commissioned by the Health Foundation, an independent charity committed to bringing about better health and health care for people in the UK. The lead authors are Jared Robert Keller, Pauline L'Hénaff and Jeni Tennison. If you want to share feedback by email or would like to get in touch, contact us at research@theodi.org.

To share feedback in the comments, highlight the relevant piece of text and click the 'Add a comment' icon on the right-hand side of the page.

We would like to thank all organisations that took part in interviews and workshops to help us put together this document.



How can it be improved? We welcome suggestions from the community in the comments.

Foreword

The outbreak of the coronavirus (Covid-19) has amplified and accelerated the need for an effective technology ecosystem that benefits everyone's health. The pandemic has been accompanied by a marked increase in the use of digital technology, including introduction of remote consultation in general practice, new data flows to support the distribution of food and other essentials, and applications to support digital contact tracing. While technology can have huge benefits, these don't always occur. If poorly designed or delivered, technology can drive up costs while adding little value, distract from other priorities, or worsen social inequalities.

While innovation seems to be everywhere, solid information about the impacts of these technologies is much harder to find. Evaluation is essential to help innovators learn and improve over time, to help health care organisations identify technologies that are having a positive benefit and spread those. But robust evaluation does not always happen as the data required to support it is often locked up within organisational silos.

This report explores models of 'data stewardship' (the collection, maintenance and sharing of data) required to enable better evaluation. It argues everybody involved in technology has a shared responsibility to enable evaluation, whether that means innovators sharing data for evaluation purposes, or healthcare providers being clearer, from the outset, about what data is needed to support effective evaluation.

This report reenvisionages the role of evaluators as data stewards, who could use their positions as intermediaries to encourage stakeholders to share data, and help increase access to data for public benefit.

Healthcare systems are seeking to accelerate the adoption of digital and data-driven technologies. While medicines and drugs are tested within randomised controlled trials, the approach for digital technologies is that they are tested 'in the wild' and adapted over time in response to learning. The approach is pragmatic, and matches the way that software developers work, but it means there is increasing reliance on the post-market assessment of benefits.

Unfortunately, the system does not always encourage the necessary data sharing to take place – meaning the issues examined in this report are becoming critical.

By analysing specific use cases, and with the help of a wider range of stakeholders and experts in the field, the Open Data Institute has made recommendations for improving data stewardship to enhance evaluation.

Healthcare providers or innovators cannot tackle this challenge alone, nor can evaluators, funders or patients. We call on all parties involved to act on the recommendations in this report, and to work together to create an environment that would allow digital and data-driven technology to reach its potential in benefiting everyone's health.

Adam Steventon *Director of Data Analytics, the Health Foundation*

Executive summary

There is much uncertainty about the impact of recent developments in technology on our health, and health and care services. New technologies offer significant potential, but active, and timely, monitoring and evaluation is crucial so that we know what impact they have on people's health and the health system as a whole, and whether it is positive or negative.

The Health Foundation and the Open Data Institute have worked together to explore how to improve access to data to support the evaluation of health technologies.

We wanted to see how data generated through digital services could be brought together with health and care data so that interventions can be evaluated in a way that is trustworthy, fair and ethical.

We explored three specific use cases: digital-first primary care technologies; online misinformation and vaccine hesitancy; and patient flow automation. Challenges were identified across these use cases in terms of data access and we examined the suitability and feasibility of new models of data stewardship to address these challenges.

Key challenges identified were related to the data needed for evaluation not being collected (sometimes due to a lack of resource), as well as the data not being accessible. Access issues can be due to a variety of reasons, such as a lack of incentives to share, worries related to sensitive data, and a lack of clarity on what is permissible. When accessed, the data can also sometimes not be as useful as expected (this can be due to a lack of quality, consistency or standards).

We explored where data institutions could be beneficial, and where they should be coupled with other elements to be put in place such as the need to convene key stakeholders and explore data needs; a push for new rules related to procurement; and the development and adoption of new standards.

Recommendations

We have identified recommendations and next steps for key stakeholders such as evaluators, funders, innovators, health and care providers, and patient and practitioner groups.

- Evaluators should:
 - consider themselves data institutions and look for ways to steward data and increase access to data for public benefit
 - use their position as an intermediary to encourage stakeholders to share data, and enable and support them to improve their capability and trustworthiness
 - act as convenors in the sector to create standards for benchmarking technologies

- Funders should:
 - explore new ways of increasing access to data through scoping and piloting new models of data stewardship such as data institutions.
 - explore other use cases in the sector such as precision medicine or the impact of social media on mental health

- Innovators should:
 - get evaluators in the room early to arrange the data collection
 - be prepared to share data for research and evaluation purposes
 - explore best practices around collecting sensitive data about who uses digital services
 - work together to develop standards for benchmarking and comparison

- Health and care providers should:
 - convene innovators and healthcare practitioners to align understanding related to data collection for evaluation purposes
 - build in evaluation from the start when piloting or deploying new health technologies
 - clarify the ways the data will be collected, accessed, used and shared at the procurement stage and in some cases embed mandatory access to data for evaluation purposes in procurement contracts.

- Patient and practitioner groups should:
 - explore cooperative models for collecting data about their experience.

Introduction

In recent years there has been a dramatic increase in the pace of innovation and the use of data and technology in health and care services. The impact of these technologies on health and care (in terms of delivery, experience and outcomes) is significant. Effective monitoring and evaluation of such technologies is critical to support confident innovation at pace.

However, there are a number of challenges related to evaluation.

The drive to innovate at pace and to avoid delays can mean new technologies are adopted before their impacts have been fully established.

The technologies themselves are largely developed by the private sector and adopted by the NHS. This means evaluation needs to combine public and private sector data, and there are both technical and governance challenges in brokering access to that data.

The evaluation of the impact of these technologies requires access to sensitive personal data. This has to be done in trustworthy and trusted ways.¹ Existing frameworks and structures do not always enable this.

There are also some blind spots in data collection, making evaluation difficult. For example, data about some parts of the system – such as social care² – might be missing or be of poor quality. Some data related to outcomes for patients (quality of life, satisfaction) is also not consistently collected.

The impact of technologies and other interventions can be different in different communities, as there are varying risks and levels of adoption due to protected characteristics such as age and ethnicity,³ or levels of digital inclusion. The Open Data Institute (ODI) has previously explored how the protected characteristics of people using a digital service can be collected, to understand how they might be affecting excluded communities.⁴

The current Covid-19 crisis is also disrupting demands on the healthcare system and the use that is made of some technologies. The pandemic will influence the future shape of health and care services in ways we cannot predict.⁵ Evaluations need to take account of this changing context.

The Health Foundation and the ODI have worked together on this project to explore how to improve access to data to support the evaluation of health technologies. We wanted to see how data generated through digital services could be brought together with health and care data so that interventions can be evaluated in a way that is trustworthy, fair and ethical.

¹ Open Data Institute (2020), '[Designing trustworthy data institutions](#)'.

² Future Care Capital (2020), '[A new joint project on data and analytics about social care](#)'.

³ Public Health England (2020), '[COVID-19: review of disparities in risks and outcomes](#)'.

⁴ Open Data Institute (2020), '[Monitoring Equality in Digital Public Services \(report\)](#)'.

⁵ The Health Foundation (2020), '[Four key questions on COVID-19](#)'.

In particular, we have examined the suitability and feasibility of new models of data stewardship. These models have the potential to bring benefits to both the public and the private sector, by facilitating and incentivising data exchange and access, while also benefiting the public, both in terms of better data governance and through the adoption of well-evidenced innovations.

Background

In this section we describe the role of evaluation in the health sector, and some of the existing work on new models for data access.

What do we mean by evaluation?

An evaluation can be defined as the process of determining the merit, worth or value of something.⁶ An evaluation of an intervention, such as the use of a health technology, can be designed to help form or shape the design or implementation of that technology, or address particular questions about its impact. The design of the evaluation needs to reflect the nature of the intervention being evaluated and the specific constraints of time, expertise and resources – including people, finance and data – available for the evaluation itself. These and other factors are used by an evaluator to determine which evaluation methodology to use, which in turn determines which questions can be answered and to what level of confidence.

A framework underpinning many evaluations is the ‘trident’, or three-pronged, approach.⁷ In this approach, evaluators seek to: measure the outcomes; describe and analyse the process; and sample multiple stakeholder perspectives. This involves addressing the questions: did it work, what happened and what did stakeholders think? Measuring the **outcomes** of technology in health and care involves assessing quality and safety, effectiveness of care and the impact on the efficiency of the system as a whole. Describing and analysing the **process** involves assessing how the technology was or is being implemented, and understanding the context in which it is used. Sampling multiple stakeholder **perspectives** involves examining the experience of patients and gathering feedback from health and care staff about their use and views of the technology.

Robust evaluations help identify whether an intervention worked, why and how. This allows lessons to be learned, existing interventions to be improved, successful interventions to be spread and scaled, and new ones developed.⁸

Evaluations of health technologies also help identify if and when a solution might be adopted more widely, for example at a national level. They also help the provider to adapt the solution being offered, based on recommendations identified.

Evidence standards are in place to help understand what good levels of evidence for digital healthcare technologies look like, and to ensure new technologies are clinically effective and offer economic value. These standards look at ways to assess aspects such as credibility, effectiveness, equality and acceptability.⁹

⁶ The Health Foundation (2015), ‘[Evaluation: What to consider](#)’.

⁷ Ellis R, Hogard E (2006), ‘[The Trident: A Three-Pronged Method for Evaluating Clinical, Social and Educational Innovations](#)’.

⁸ Ibid.

⁹ National Institute for Health and Care Excellence, ‘[Evidence standards framework for digital health technologies](#)’.

A code of conduct for data-driven health and care technologies has also been developed by NHSX, to proactively and objectively evaluate current technologies, and ensure best practices are developed and implemented in an evidence-based manner.¹⁰

The Health Foundation has identified some key challenges for evaluators of health technologies:¹¹

- Are some patients more likely to be excluded because of their complex needs or experience using digital services?
- As technologies evolve through self-learning, how often do they need to be evaluated, and how reliable can the findings be over time?
- How can we ensure that the understanding of how data is collected and used is common across all evaluators, and build this capability?
- How can we ensure the evaluation considers the context in which technologies are introduced, so that we can understand what might be contributing to their impact?

There is also a key challenge – common to all evaluations – around collecting the right data and making it accessible to evaluators. Evaluations can be thought of late in the process, which means delays in identifying the questions to be asked and the data necessary to answer them. Evaluators may not be able to access the data they need to perform a robust evaluation, particularly within the time constraints of an evaluation.

What do we mean by new models of data stewardship?

Stewarding data involves collecting, maintaining and sharing it, and as part of this, making decisions about who has access to it, for what purpose and to whose benefit. When data is stewarded responsibly, data is available to those who need it in ways that are trustworthy and sustainable.

As part of its work to build an open, trustworthy data ecosystem, the ODI has been exploring different approaches to stewarding data. Other organisations, such as [The GovLab](#), the [Aapti Institute](#) and [Nesta](#), and networks such as those run by [MyData Global](#), [Mozilla](#) and the [Centre for International Governance Innovation](#), are also exploring this topic.

At the ODI, we think data institutions have an important role to play in the stewardship of data. Data institutions are organisations whose purpose involves stewarding data on behalf of others, often towards public, educational or charitable aims. The ODI's initial work has found that data institutions play a number of vital roles in different sectors and contexts, including:

- holding data on behalf of an organisation or person, or group of them, and sharing it with others who want to use it for a particular purpose
- combining or linking data from different sources, and providing insights and other services back to those that have contributed data

¹⁰ Morley J, Joshi I (2019), '[Developing effective policy to support artificial intelligence in health and care](#)'.

¹¹ The Health Foundation (2019), '[Evaluating digital first primary care – the challenges ahead](#)'.

- maintaining common data infrastructure for a sector or field, such as by registering identifiers or publishing open standards.

There are many existing data institutions, including in the health sector. The UK Biobank is one example. It was set up in 2006 to steward genetic data and samples from around half a million people, and continues to support their use for health research. The data being stewarded by the Biobank is available to health scientists from academia and industry for research purposes, via an application process. The objective is to improve the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses.¹²

There are many different models for facilitating access to data for public benefit. In the context of this report, the need to share potentially sensitive data from both the public and private sectors, in a trustworthy way, towards the specific purpose of evaluating the impact of health technologies, meets many of the criteria for a particular form of data institution known as a ‘data trust’.¹³ However, appropriate data stewardship and access models are context sensitive. Their design is dependent on the details of a given data ecosystem and the motivations and capabilities of different stakeholders.

Adopting new models for data stewardship does not always mean creating a new data institution. Often, it may be more appropriate for an existing organisation to take on a new role or to recognise its status as a data institution. In this report, we consider both gaps where new data institutions might provide trustworthy access to data, and places where existing organisations could or should adopt this role.

¹² UK Biobank, [‘About UK Biobank’](#).

¹³ Open Data Institute (2019), [‘Data trusts: lessons from three pilots \(report\)’](#).

Three use cases

This report focuses on three use cases related to the evaluation of emerging technologies and their impact on health and care in the UK. By focusing on specific use cases and technologies, we have been able to take an in-depth look at the data ecosystems that surround these new technologies, their potential value, the relevant stakeholders involved, and the challenges and barriers to accessing and sharing data for evaluation purposes. In particular, we have been able to identify where within these three use cases there is potential for new forms of data stewardship to help remove barriers to data access and sharing.

The insights and lessons we have drawn from this research are applicable beyond these specific technologies and contexts. This report therefore offers recommendations to evaluators and other organisations working with these specific technologies, alongside overarching guidance and lessons for organisations working on emerging technologies in the health and care sector more broadly.

We chose three complementary use cases:

- Assessing the safety, effectiveness and efficiency of **digital-first primary care** services designed to advise patients about their symptoms and direct them to other, non-digital primary care services.
- Understanding the spread of **information and misinformation** online, and the impact of misinformation on **vaccine hesitancy** and public health.
- Evaluating **patient flow automation** systems which are designed to improve clinical pathways and operational efficiency within hospitals and across regions.

The use cases have contrasting and complementary characteristics, as summarised in the table below.

	Digital-first primary care	Misinformation and vaccine hesitancy	Patient flow automation
Location within health sector/society	First point of contact with the health sector	Outside the health sector but impacting health	Embedded within the health sector
Level and manner of engagement with people	Patient-facing for specific purposes	Frequent contact with people but outside the health sector	Primarily 'behind the scenes'
Extent of adoption	A handful of implementations	Large-scale adoption	Dozens of implementations
Type of technology	Specific technology used for particular purposes when	Suite of technologies used for a range of purposes	Specific technology used regularly

	needed		
Primary level/ scale of impact	Impact on individual users as well as care providers	Impact on society at large	Impact on hospitals or regions

Relevance for Covid-19

The three use cases were selected in early 2020, before social distancing measures were put in place to prevent the spread of Covid-19, but each technology has played an important role in the health sector's response to the Covid-19 pandemic.

- **Digital-first primary care:** In the weeks after social distancing measures were put in place, virtual GP consultations markedly increased.¹⁴ The need to facilitate continued access to health services has highlighted the value of virtual consultations, as well as being able to triage the types of people and symptoms who should be seen in person.
- **Misinformation and vaccine hesitancy:** Much misinformation has spread online during the pandemic,¹⁵ highlighting the importance of being able to direct people toward helpful, authoritative sources of health information and away from potentially harmful misinformation. This is likely to take on even greater importance if/when a Covid-19 vaccine is produced.
- **Patient flow automation:** During the current pandemic, hospitals and the health systems have been stretched thinly. The need to use available resources effectively and efficiently has highlighted the value of having a real-time view of resources and capacity across a system, and the ability to model and anticipate peaks of activity and potential bottlenecks.

We have been able to observe the impact of the pandemic on such technologies and on efforts to share and access data more broadly. New temporary data-sharing procedures and agreements have been put in place at a rate that would have been difficult in normal circumstances. For each use case, we draw out important lessons learned from the pandemic and provide recommendations and next steps related to efforts to battle Covid-19.

The following three sections of this report examine these use cases in more detail, based on our desk research, interviews and workshops with experts and stakeholders. We provide background on the specific technology, the context and the results of our research. Specifically, we outline:

- the types of things an evaluator would look to assess or demonstrate through an evaluation of that technology

¹⁴ Financial Times (2020), '[Lockdown drives boom in healthcare apps](#)'.

¹⁵ BBC (2020), '[Coronavirus: Fake news crackdown by UK government](#)'.

- the types of data that an evaluator would need access to in order to perform that evaluation
- the challenges or barriers that currently exist to accessing those types of data.

At the end of each section we outline recommendations and next steps, including indicating where some of the challenges and barriers identified can be addressed by a data institution. Where appropriate, we also note where challenges and barriers can be addressed through other interventions, including non-institutional forms of data stewardship.

Many of our findings and recommendations apply more broadly to multiple use cases or to the health sector in general. We outline these overarching findings and recommendations after discussing each of the three use cases.

Use case #1: Digital-first primary care

Use case description and background

Digital-first primary care (DFPC) refers to models of general practice where patients use smartphone and desktop applications to book and conduct consultations remotely.¹⁶ These digital and online tools make it easier for some patients to access the advice, support and treatment they need.

There are a number of approved DFPC providers, offering different models, including triage only, triage with consultation (either video or online) and full digital primary care.¹⁷ Examples of such services include [Babylon GP at hand](#), [eConsult](#) and [Ada Health](#).

The NHS Long Term Plan¹⁸ commits that every patient will have the right to DFPC by 2023/24, with online consultations by April 2020 and video consultations by April 2021.¹⁹

However, there have been relatively few formal evaluations of the impact of DFPC in general practice – on patients, the workforce and the wider health and care system. For example, it is currently difficult to assess whether the decisions made when using a DFPC service are causing added strain on a particular section of the health and care system, or whether a DFPC provider is creating a digital divide.²⁰

Findings

This section contains a summary of our findings. For further details see the [research findings appendix](#).

What an evaluation would aim to assess or demonstrate

Based on our research, an evaluation of DFPC technologies would pursue a three-pronged evaluation, similar to many other evaluations of emerging health technologies.

As outlined in the [Background section](#), within such an approach, evaluators would seek to measure the outcome of the introduction of a DFPC technology (what are the patients' experiences and pathways, and their views on the quality of service?); describe and analyse the process by which the technology was introduced (does the service create inequalities in terms of patients access?); and sample multiple

¹⁶ The King's Fund (2019), '[Digital-first primary care: helpful disruptor or unnecessary disruption?](#)'.

¹⁷ NHS England, '[Digital First Primary Care](#)'.

¹⁸ NHS England (2019), '[NHS Long Term Plan](#)'.

¹⁹ NHS England, '[Digital First Primary Care](#)'.

²⁰ Ibid.

stakeholder perspectives on the use and value of the technology (what are the practitioners' experiences using it? Does the service provide value to the healthcare system and make it more efficient?).

Data an evaluation would need

In order to assess or demonstrate the impact of a DFPC technology, evaluators would need to be able to access and link a wide range of data:

- Data from general practitioners (GPs) and other health and care staff related to their experience of using DFPC technologies.
- Data about patients (including demographic data) and how they use and interact with DFPC technologies, how they are directed through the health system after they have used it, and what their experience of using the technology was.
- Operational data drawn from commercial providers as well as from health providers. This data would include aggregated usage data related to how and when the tool was used, NHS system-level data about waiting times and GP workflows, and data about any decision-making algorithms used within those technologies and any data used to train those algorithms.

Challenges or barriers to accessing necessary data

According to the stakeholders and innovators we interviewed within this area, it can often prove difficult to access necessary data..

Sometimes, the data needed to perform an evaluation is simply not collected. This can be due to a lack of resources (time and capacity), high data-collection costs, or because the need to collect a specific type of data was not identified or communicated from the start.

Other times, the data exists or has been collected, but is not accessible to the extent desired by evaluators; whether due to concerns about commercial sensitivity, or user or patient privacy. This barrier to accessing data appeared frequently in our research on DFPC technologies.

In some cases, the necessary data has been collected and is accessible, but is not as useful as desired. According to our research, this is often because the data is not granular or real-time enough, is of poor quality or is anonymised to the point that it is less useful for the purposes of evaluations. At times, the necessary data lacks consistency or shared standards, making it difficult to link disparate datasets or perform analyses across aggregated datasets.

Recommendations and next steps

The two main challenges and barriers identified within this use case are gaining access to data held by commercial innovators and technology providers; and accessing necessary data from across the NHS and the health sector. Here, we discuss these barriers, offer recommendations for addressing them, and then outline suggested next steps.

Embed mandatory access to data for evaluation purposes in procurement contracts

The people and organisations we spoke to identified several specific commercial datasets and types of data that are currently difficult to access, including data about people's use of DFPC tools and services; the algorithms and models used to make decisions about recommended treatments or where to direct patients; and data used to train algorithms and models, if applicable.

The main barriers to accessing commercial data are a potential lack of trust between commercial providers and evaluators (though not in all cases); and a perceived inability to incentivise commercial organisations to share potentially sensitive data, because it contains either personally identifiable or commercially sensitive information.

Building or empowering a data institution would help to address some of the current barriers to accessing commercial data, but would not be sufficient on its own. It would be possible to build or install a data institution to sit between evaluators and commercial technology providers to balance competing interests, act as a trusted third party and ensure the security of personally and commercially sensitive data. But a data institution on its own would not be able to address the inability to incentivise or compel commercial organisations to share data with evaluators. For this, another approach is necessary.

In this use case, the data holders – ie the innovators and DFPC providers – want access to the NHS market and the potentially lucrative contracts that may come from having their technology or solution adopted by healthcare providers. Having their technologies evaluated is a crucial part of that process. Therefore, evaluators and NHS bodies do have a degree of leverage. Drawing on that leverage, it should be possible to include specific requirements in contracts with DFPC providers that oblige them to share data with evaluators for the purpose of evaluation. Evaluators will need to work with NHS organisations to draw up terms that are amenable to all parties and to ensure that these obligations are added to contracts early in the process of implementing, piloting and procuring new technologies.

Such an approach would enable evaluators to gain access to data currently held by innovators and commercial technology providers. Any data institution is unlikely to succeed without it.

Enable institutional efforts to improve access to NHS data to support technology evaluation

Another challenge we identified in this use case (and others) is that **it is difficult to gain access to necessary data from across the NHS and the health sector more broadly.** In some cases, this data is not collected; in others, the data is collected but is not available in the desired form or is not linkable with other sources. The people and organisations we spoke to identified various reasons for this, including but not limited to:

- a lack of sufficient technical infrastructure
- a lack of consistent and suitable standards
- poor data quality
- concerns about patient privacy.

There is a role for data institutions to play in increasing access to data across the NHS and the health sector, and there are initiatives already underway that aim to address many of these challenges and barriers.

For instance, Health Data Research UK (HDRUK) is currently working to unite, improve and use data from across the health sector for research purposes through the [UK Health Data Research Alliance](#) and the [Health Data Research Innovation Gateway](#), and to make health data more accessible and useful through the establishment of a series of [Health Data Research Hubs](#).²¹ HDRUK is also working to ensure that health data is put to good use through funding regional and national innovation programmes with organisations such as the National Institute for Health and Care Excellence (NICE) and the Health Foundation.²²

NHS Digital, which is responsible for standardising, collecting and publishing data from across the health sector in England, is also conducting work that seeks to address many of the challenges and barriers identified.²³ Its work aims to develop national and international standards for health data, improve the quality of health data, and increase the accessibility of health data.²⁴

These two organisations are not focused specifically on bringing together data for the purposes of evaluations of digital health technologies, but they are working to address many of the same barriers to access that currently confront evaluators. Ongoing initiatives by these and other organisations within the health sector should engage with evaluators to ensure that they can get timely access to data for evaluation purposes. For instance, in an interview with HDRUK, we discussed ways that one of the research hubs, Discover Now, may be able to facilitate access to data for the purposes of evaluations.

Engage with existing initiatives and consider piloting a data institution to facilitate access to NHS data for the purpose of evaluations

It is currently unclear whether existing initiatives to increase access to health sector data, such as those outlined above, will meet the needs of evaluators, although it should still be possible to learn from them.

If these central initiatives do not meet the needs of evaluators – for instance, it may be that these approaches will not be agile enough to meet the needs of evaluations that need to move at pace – then evaluators should **explore the potential of taking on the role of a data institution to facilitate access to data in a way that does meet their needs**. For example, this institution could facilitate access to data held by individual NHS Trusts or clinical commissioning groups rather than going via the more national bodies. We describe what such a data institution might look like and provide potential next steps for evaluators in use case #3 and in the conclusions.

²¹ Two relevant examples are the [‘Health Data Research Hub for Real World Evidence’](#) and the [‘Health Data Research Hub for Clinical Trials’](#).

²² Health Data Research UK, [‘How we use health data’](#).

²³ NHS Digital, [‘Data and information’](#).

²⁴ NHS Digital, [‘Data, insights and statistics’](#).

Explore the potential for existing organisations to take on new data stewardship roles

We believe there are several important data stewardship roles that existing organisations working in this space should and could perform. Here we draw a distinction between *creating* new institutions to perform specific roles and *empowering existing organisations* to take on new roles or to recognise their status as a data institution. Performing these roles well will help increase access to health data for the purpose of evaluating DFPC technologies, while also increasing access to health data more broadly.

Through our use case workshops and interviews with organisations such as HDRUK, the National Data Guardian and NHSX, we identified areas where further work is needed. An organisation wishing to facilitate access to data for evaluation purposes within this space could action the following points:

- Push for the development and adoption of new procurement terms and conditions for services that include the requirement to provide access to data and other materials for the purpose of evaluating the performance of the service.
- Push for the development and adoption of new standards, certifications or accreditations of organisations involved in the collection, processing, sharing and use of health data. This can help organisations build and demonstrate their own trustworthiness, while helping them assess the trustworthiness of others.²⁵
- Push for new policies or regulations where necessary, for instance, policies that address challenges related to the legal basis for accessing patient data.
- Help provide clarity for organisations around what is and isn't permissible in terms of the collection, use, linking and analysis of potentially sensitive datasets.
- Develop a framework to enable consent-based access to data across commercial and NHS organisations for evaluation purposes. Some evaluations will require access to many different types of data from many different sources, but some can be performed with less data from fewer sources. In these cases, a consent-based approach may be workable and managing those consents could be a role taken on by an evaluator acting as a data institution.
- Engage with commercial organisations to advocate the benefits of sharing or opening data for the purposes of evaluations. Though we encountered some resistance from commercial organisations to sharing certain types of sensitive data, it may still be possible in some circumstances to persuade commercial providers to increase access to data they hold, especially if the public health and societal benefits are made apparent.

Advocate for technologies to be implemented in a more evidence-based way, with early input from evaluators

There should also be advocacy for emerging health technologies to be implemented in a more evidence-based way. This advocacy could be led by an evaluator or by an NHS body, but regardless, this type of initiative would likely require contributions and cooperation from numerous organisations inside and outside the health sector.

²⁵ Open Data Institute (2020), '[Designing trustworthy data institutions](#)'.

Approaching the implementation of emerging health technologies in a more empirical manner would enable evaluations to be conducted in a more rigorous, controlled, scientific way than is currently possible. Evaluators are often consulted or contracted *after* a new technology has already been rolled out, meaning it is difficult – sometimes impossible – to collect the types of data necessary or control for important variables. Including evaluators in discussions early on in the roll out of a new technology would make it possible to identify important things to evaluate, important data to collect, and potential challenges or barriers to collecting that data. This would also enable evaluators to communicate to health personnel the necessity and value of collecting that data, which should increase buy-in.

By working with the NHS to implement emerging health technologies in this way, evaluators can help the NHS to continue to innovate at pace, while making it easier to evaluate the safety, efficiency and effectiveness of these technologies.

Covid-19 context

The uptake and roll out of digital-first primary care solutions has increased massively due to Covid-19. The pandemic has resulted in a move towards delivering a ‘total triage’ model of care through online, video and telephone consultations. In May 2020, online consultations were available in 85% of general practices, covering 86% of the population; and video consultations were available in 99% of practices, covering 99% of the population²⁶ (while in 2019 less than one in every 100 of all GP appointments was carried out by online video consultation).²⁷

Some questions, however, still remain in terms of:

- understanding which parts of health care services are still better addressed in person and which can best be conducted digitally.
- understanding the impact these technologies can have on some groups that are excluded (creating a digital divide). Some technologies were identified as being used mostly by younger and healthier people, while older people with more complex health needs were less frequent users.²⁸

Evaluating these issues would be necessary before adopting such technologies further and fully transitioning to digital care in a post-Covid-19 context.

²⁶ Bakhai M, NHSX (2020), ‘[The use of online and video consultations during the COVID-19 pandemic - delivering the best care to patients](#)’.

²⁷ The guardian (2020), ‘[GPs told to switch to digital consultations to combat Covid-19](#)’

²⁸ Ipsos Mori (2019), ‘[Evaluation of Babylon GP at hand](#)’.

Use case #2: Online misinformation and vaccine hesitancy

Use case description and background

There are growing concerns about the impact of the online spread of information and misinformation on many areas of life, several of which might impact health. For this use case, we looked at the impact of online information and misinformation on vaccine hesitancy.

Online information includes that found on social media websites and applications where users can create or share content. It also includes other sources of information online. For this use case, we also considered messaging applications (that is misinformation might be flowing quickly through messaging applications rather than on platforms). The information may or may not be accurate, and inaccuracies may be intentional or unintentional.

Hesitancy in relation to vaccination may affect people's motivation to be vaccinated, causing people to reject it for themselves or their children.²⁹ This hesitancy can be caused by a number of factors, including complacency about the need for vaccinations, difficulties with access to and lack of confidence in vaccinations. The World Health Organization identified vaccine hesitancy as one of the top 10 global health threats of 2019.³⁰

Misinformation around vaccinations is a long-standing problem, however social media presents unprecedented risks around the amplification and spread of anti-vaccination messages.³¹

Findings

This section contains a summary of our findings. For further details see the [research findings appendix](#).

What an evaluation would aim to assess or demonstrate

Evaluating the impact of online information and misinformation on vaccine hesitancy is a difficult and complex task. This is in large part due to the diffused but widely

²⁹ World Health Organization (2019), '[Improving vaccination demand and addressing hesitancy](#)'.

³⁰ World Health Organization (2019), '[Ten threats to global health in 2019](#)'.

³¹ Burki T (2019), '[Vaccine misinformation and social media](#)'.

adopted nature of the technology being evaluated. Whereas the previous use case focused on a fairly discrete type of technology in the early stages of adoption, this use case focuses on a suite of related but different technologies that are widely used and deeply embedded within many aspects of people's lives. Because of this, the three-pronged approach to evaluations used in the previous use case would be insufficient.

To assess and demonstrate the impact of online information and misinformation on vaccine hesitancy, evaluators would need to, among other things: identify and measure the types of misinformation about vaccines circulating online; understand the motivations and methods of people propagating misinformation about vaccines; understand the role that online platforms play in either limiting or circulating misinformation; track the types of misinformation about vaccines people have been exposed to; understand what factors influence a person's susceptibility to misinformation; understand the influence of offline sources of information in shaping a person's views of vaccines; and ultimately be able to measure the extent to which exposure to online content influences people's actions and behaviours. This last step is crucial to proving causality rather than correlation.

Data an evaluation would need

Because the answers sought by evaluators are complex and multi-faceted, a comprehensive evaluation will require access to many different types of data from many different sources.

Evaluators would need data from sources within the health sector related to visits to GP practices and vaccine clinics, data about messaging and communication about vaccines, and data about the overall rates of vaccine uptake across the country. Evaluators would also need contextual data about people's demographics, socioeconomic backgrounds, beliefs and offline activities.

Crucially, evaluators would need data from online platforms and services, including data about the types of vaccine information and misinformation available online, and data about how people view, share and comment on misinformation. Evaluators would also need data related to any algorithms trained to prioritise and circulate content online.

Challenges or barriers to accessing necessary data

Based on our discussions with stakeholders and innovators within this area, there are common challenges and barriers to accessing the data needed to assess the impact of online misinformation on vaccine hesitancy.

As with the previous use case, sometimes the necessary data is not collected due to a lack of resources, concerns about privacy or because the data was not identified as important for evaluations and was therefore not collected. There is also often a lack of clarity about whether collecting some types of data is legally permissible.

When data does exist, it can be difficult to access – especially data that is held by online platforms and commercial tech companies. Often, these companies are unwilling to share data due to concerns about commercial sensitivity and intellectual property, protecting the privacy of their users, and the ethics of sharing data about

users. Some of our interviewees noted that because the data is of a personally or commercially sensitive nature, it is very difficult to find a way of incentivising organisations to share that data.

Even when it is possible to access data from online sources, a lack of consistency or standardisation across platforms for flagging, labelling, reporting and removing misinformation tends to make it difficult to perform cross-platform analyses.

Recommendations and next steps

Not surprisingly, we identified many of the same challenges in this use case as in the previous use case around digital-first primary care. In some cases, the particulars of this use case have led us to outline different recommendations or add nuances to the recommendations outlined in use case #1. In others, our recommendations remain largely the same, in which case we refer back to the previous use case rather than repeating our recommendations here.

The challenge of accessing and linking data from across the NHS and the health sector came up often within this use case. For more on our recommendations for addressing this challenge, see [‘Use case #1: Recommendations’](#).

Explore the use of data portability to collect contextual data from offline sources

An important challenge in this use case is that **it is difficult to gain access to contextual data and data about people**. A few of our interviewees and workshop attendees noted the potential for evaluators and health organisations to conduct surveys to collect demographic details, data about people’s activities offline and attitudes toward vaccines. This is a worthwhile avenue to pursue and sits firmly within the standard research methodologies of social scientists and public health researchers.

Another potential avenue would be to develop ways for people to participate in ‘data altruism’: leverage their right to data portability to provide evaluators with access to important types of data related to misinformation and vaccine hesitancy. Data portability allows individuals to obtain and reuse personal data about them for their own purposes across different services (they can move, copy or transfer personal data easily from one IT environment to another in a safe and secure way, without affecting its usability).³² People could port relevant records from health systems, local authorities, social services or schools to provide evaluators with demographic details or details about a person’s interactions with various offline services. This type of information would help evaluators gain a clearer understanding of the wide range of different factors that influence vaccine hesitancy and might even help health and care officials target interventions.

Data portability could also be used to gain access to data held by online sources, but there are additional potential limitations in this case. For example, there are limits to the type of data a person can request from an online platform. In particular, a user would only be able to receive data that they themselves created or contributed to the

³² Information Commissioner’s Office, [‘Right to data portability’](#).

service, for example, data about the posts *they* have made, but would not be able to receive data about all the posts they had seen, or about the advertising directed at them. Nor would a user be able to receive access to algorithmic decision-making related to *why* certain posts or adverts were directed at them.

A data portability initiative, whether aimed at gathering data from offline or online sources, would probably rely on people volunteering to take part. Whether it would be hindered by the self-selected nature of the cohort who sign up would need to be explored. It is unclear, for instance, whether the people most likely to reject vaccine advice would be the people signing up for this type of data portability scheme. This touches on a wider point about ‘crowdsourced’ or ‘patient-led’ initiatives: though they have proved useful as a means of increasing access to data for a range of research purposes, there are questions as to how reliable, representative and timely they can be expected to be and therefore whether evaluators should make them a core part of their evaluations.

However, as Gary King of [Social Science One](#) has said, researchers are adept at using facts we know to learn about facts we do not know. So, despite the potential limitations of such an approach, we recommend exploring the potential of data portability further. In a use case such as this – as in many others – every little bit of data is likely to help.

Explore the use of plugins and data portability to collect data from online sources

Another challenge is that **it is difficult to gain access to data from online sources, in particular from online platforms and social media companies.** There is a perceived inability to incentivise or compel social media companies to provide access to this data, as well as concerns over commercially sensitive data and user privacy. As such, the challenges are similar to the challenges identified in the first use case, but with an important differentiating factor. In use case #1, evaluators and the NHS had leverage in the form of access to the NHS market which they could use to incentivise digital-first primary care providers to share necessary data. In this second use case, since online platforms exist outside the health sector and do not rely on NHS contracts, evaluators do not have the same degree of leverage or ability to incentivise cooperation. Therefore, whereas in use case #1 we recommended pursuing contractual access to data, in this use case we recommend exploring a few different approaches.

One approach would be to explore data portability in order to access data held by online platforms, but there are limitations to this approach, as outlined above.

A related approach would be to explore the potential of browser plugins to gather information about people’s interactions with online services. For instance, a [browser plugin developed by ProPublica](#) enabled people to “see exactly how Facebook users are being targeted by advertisers”.³³ These types of tools would not give evaluators a view into algorithmic decision-making, but they would enable them to track and catalogue adverts and see which groups advertisers are targeting with which advertisements. The legality of these tools is currently contested, however, so organisations should explore these with caution. For instance, Facebook “urged” ProPublica to shut down their plugin.

³³ ProPublica (2019), ‘[Facebook Moves to Block Ad Transparency Tools - Including Ours](#)’.

In both these cases, the data collected could be held by a data institution that would then provide access to that data to a range of researchers and evaluators.

Consider piloting a data institution aimed at facilitating access to data held by online platforms

Given the necessity of accessing data held by online platforms and the limitations of the two approaches outlined above, we think that **a data institution could play an important role in facilitating access to data held by online platforms.**

Our research into similar data institutions suggests that a data institution aimed at facilitating access to data held by online platforms would be able to incentivise these platforms, while addressing concerns about commercial sensitivity and user privacy.

In terms of incentives, collaborating with a data institution set up to study the impact of misinformation on vaccine hesitancy would provide online platforms with:

- **An opportunity to share data for public benefit.** Many online platforms already share data for such purposes, so an initiative aimed at helping to evaluate the impact of online misinformation on vaccine hesitancy and public health would fit firmly within those existing initiatives and motivations.
- **A chance to garner good public relations.** Helping to answer such an important question would give online platforms a chance to show their commitment to promoting public health, while also potentially countering any criticism of the way that misinformation spreads on their platforms.
- **An opportunity to forestall regulation.** Some data institutions and data access initiatives like Uber Movement have arguably been set up as a way of heading off regulated access to commercial data that might have been on terms less generous to the commercial data holder.

An independent data institution would be better placed to address concerns about user privacy and commercial sensitivity than other actors. By sitting between private, public and third sector organisations, data institutions can serve as trusted third parties and help to increase access to data for research purposes, while ensuring that nothing commercially or personally sensitive is accessed by the wrong parties. By utilising tools and approaches like [differential privacy](#) or secure research environments, a data institution could help commercial data holders – and the people that use these platforms – have confidence while sharing data.

This data institution could take a form similar to [Social Science One](#), the industry-academic partnership set up to explore “the effect of social media on democracy and elections”. Or it could resemble other data-sharing initiatives involving major data-holding companies such as Facebook’s [Ad Library](#), [SharedStreets](#) or [Uber Movement](#).

Before the exact structure and focus of this potential data institution can be decided on, however, a number of questions about the remit and breadth of the institution will need to be answered. Most of these questions are a consequence of the fact that the answers sought by evaluators in this use case are quite complex (proving causality versus correlation is a notoriously difficult task) and will require access to many different types of data from many different sources. These open questions include:

- **Should this data institution focus on a single health topic like vaccine hesitancy or on public health more generally?** Social Science One, as a comparison, is aimed broadly at questions of democracy and elections, but there may also be value in starting with a smaller focus and expanding from there.
- **Should this data institution start by aiming to facilitate access to one online platform or to many?** It is unclear whether online platforms would be willing – or able – to cooperate with each other on such an endeavour.
- **Should this data institution exclusively focus on facilitating access to data from online sources?** Or should it set its sights on facilitating access to data from all the different sources necessary to perform an evaluation of this case, that is data from the health sector, contextual data, data from people, academic research data, and so on.

Engage with stakeholders to identify their views and needs

We recommend conducting a scoping process to confirm the need for a data institution and explore what such an institution would look like. Our report, '[Data trusts: lessons from three pilots](#)', has resources that can help evaluators in this effort.

³⁴ In particular, our recommendation for evaluators is continued research and stakeholder engagement on three fronts:

- **Convene a range of online platforms and companies to discuss their views in this area.**
 - What are their views on data sharing and cross-industry cooperation?
 - What type of data institution would be agreeable and what types of incentives could convince them to take part?
- **Engage with researchers – be they social scientists, evaluators or epidemiologists – to discuss their needs around this use case in more depth.**
 - What types of questions do they want to answer and what types of data would they need in order to do so?
 - In what form would they need this data? Would differential privacy or a secure research environment suffice?
 - What types of questions would be answerable if the data institution only facilitated access to data from online platforms? What if all the necessary types of data were made available?
- **Engage with similar initiatives exploring similar terrain.**
 - For instance, research into [online harms](#) or calls for regulated access to social media data in order to enable research into the impact of [social media on mental health](#).

Rethink the scope of evaluations on this topic

A final recommendation is for evaluators and public health researchers to **focus on using available data to answer smaller aspects of the larger question**, rather than aiming for a single source of all possible relevant data. Answering the ultimate question of the impact of misinformation on vaccine hesitancy will require access to

³⁴ Open Data Institute (2019), '[Data trusts: lessons from three pilots](#)'.

many different types of data that are currently not available, but in our workshops and interviews we discussed the potential of using the few bits of relevant data that are currently available to begin answering smaller questions related to vaccine hesitancy and misinformation, as is the standard scientific process. Once researchers have answered a number of smaller questions in this area, the answers can be pieced together to provide a view of the larger question.

These individual studies are useful not only in shining a light on the research questions themselves, but also in highlighting the utility of certain datasets and therefore prioritising investment. By examining existing and recent research, evaluators should be able to identify which datasets from which sources have proved the most useful in order to prioritise investment and effort aimed at increasing access to those datasets.

Covid-19 context

In the context of the Covid-19 pandemic, vaccine hesitancy remains an important issue,³⁵ both in terms of maintaining scheduled vaccinations during periods of social distancing, but also to ensure access to high quality and accurate information as Covid-19 vaccination research – and hopefully eventual deployment – is underway. With reduced access to health professionals, this challenge is even more present as people might rely more on other sources of information.

There is a form of easy-to-reach consensus around the need to fight the spread of Covid-19 worldwide. In this context, people may have fewer privacy concerns when the sharing of data about them is for uses that the public think are important. Platforms might be more easily pushed (both by the public and internally) towards taking an active role in the fight against the virus.

As research begins on potential vaccines for Covid-19, a consensus might be harder to reach in the future. Opinions might differ as to how best to treat the virus, and some might strongly oppose vaccination. This might make it more difficult for people to agree on the use of data about them, and platforms might be more cautious about playing an active role.

³⁵ The Colonel Group (2020), '[A future vaccination campaign against COVID-19 at risk of vaccine hesitancy and politicisation](#)'.

Use case #3: Patient flow automation

Use case description and background

Improving the flow of patients, service users, information and resources within and between health and social care organisations can play a crucial role in coordinating care around the needs of patients and service users, and driving up service quality and productivity.³⁶

Poor patient flow is not only a source of significant waste and delay, but it can also be devastating for patients and service users, and deeply frustrating for people working in health and social care.³⁷ Poor patient flow contributes to crowded and unsafe emergency departments; patients being admitted to wards that are not best suited for managing their care; poor clinical outcomes (especially for frail patients); poor patient experience (for example being moved between wards or delayed care); and poor staff satisfaction.

Population health management³⁸ has highlighted the importance of understanding how patients flow through different pathways and between different care providers across the health and care system.

Patient flow management tools can help with making these pathways more efficient and can lead to better health outcomes. They can help map flows, and so enable analytics and modelling, for example:

- real-time views of patient pathways and capacity within hospital wards for bed management
- opportunity to link in information from others (for example, primary care or community) and to share information on care records
- opportunity to model patient pathways to help understand patient flow and perform 'what-if' analyses to identify more effective and efficient service configurations.

Patient flow management tools are developed both 'in-house' by NHS analysts and by external organisations. They allow better reporting and enable managers and clinicians to access information closer to where the decisions get made.³⁹

Examples of such tools include the patient tracker⁴⁰ and the operational control centre⁴¹ developed by [Beautiful Information](#) (helping to see performance or capacity

³⁶ The Health Foundation (2016), '[The challenge and potential of whole system flow](#)'.

³⁷ Ibid.

³⁸ NHS England, '[Population Health and the Population Health Management Programme](#)'.

³⁹ The Health Foundation (2019), '[Untapped potential: Investing in health and care data analytics](#)'.

⁴⁰ Beautiful Information, '[Patient tracker live](#)'.

⁴¹ Beautiful Information, '[Performance](#)'.

within a hospital, in real-time) and [Clinithink](#) (helping to manage patient pathways and improve resource management).

Findings

This section contains a summary of our findings. For further details see the [research findings appendix](#).

What an evaluation would aim to assess or demonstrate

Based on our research, an evaluation of patient flow automation tools would follow an approach similar to the one outlined in use case #1; focusing specifically on assessing the impact of these tools on patients and the health system as a whole. Evaluators would seek to measure patient experience, patient outcomes and any unintended consequences in the treatment that patients receive.

In terms of the impact on the whole system, evaluators would seek to assess the impact that an implementation within a hospital might have on social care services, GP practices and community care services; focusing on whether the tool is causing added strain on a particular part of the system.

Data an evaluation would need

To assess or demonstrate all of the above, evaluators would need to be able to access and link data from a number of sources across local and regional health systems. They would need data on how and when a specific patient flow automation tool was used and any actions taken as a result of using that tool. They would need data from across the system on where patients were directed and what capacity was when they were directed there. Evaluators would also need data about any decision-making algorithms used within these technologies and access to any data used to train those algorithms.

Finally, they would need data collected from patients related to their experience and overall satisfaction.

Challenges or barriers to accessing necessary data

Our research identified several challenges and barriers to accessing necessary data, many of which are similar to those identified in the previous two use cases. As in the first use case, we identified challenges related to accessing, linking and analysing data from across the health sector, as well as challenges related to accessing data held by commercial innovators, often due to concerns about personal and commercial sensitivity.

Similar to use case #2, there can often be a lack of clarity around what is legally permissible when it comes to the collection, sharing, linking and use of certain types of data. This lack of clarity can extend to control of data as well. Several people we spoke to noted that it is often difficult to know who controls different datasets and

therefore who to approach to request access, especially when a hospital or health system is using digital technologies from multiple different providers.

An additional challenge raised in many of the interviews was that some teams within hospitals are hesitant to share data because they are concerned that it will be misunderstood by other departments and parties, and that misunderstandings could come back to impact them in a negative way.

Recommendations and next steps

The main challenge with this use case is **gathering the data necessary to evaluate and compare numerous implementations of similar technologies**. This is in large part a consequence of the fact that patient flow automation tools and services are the most developed and widely adopted of the health technologies examined for this project. Though digital-first primary care tools are gaining wider adoption – and are likely to be implemented more widely following their use during the Covid-19 pandemic – patient flow automation tools are already embedded in NHS processes.

Consider piloting a data institution aimed at enabling comparative and long-term evaluations

The systems currently in use in the health sector include some that have been developed by teams within NHS hospitals and trusts, and some that have been developed by commercial companies. For evaluators, the difficulty in evaluating and comparing these stems from the fact that each implementation is bespoke with different technical setups, standards and legal agreements with data sources. The bespoke nature of the implementations requires evaluators to work in a similarly bespoke manner in order to collect, analyse and work out the legal agreements to access necessary data. Comparative evaluations are therefore possible, but require large amounts of time, effort and resources.

There is a potential role for a data institution to play here. In short, this data institution could work with public and private sector innovators to collect and steward relevant data for the purposes of evaluating patient flow automation tools; while in return, the data institution could provide innovators with aggregated insights such as benchmarking and safety reports that help them improve their services.

Through such an institution, evaluators would receive access to data about multiple patient flow automation tools, which would help them evaluate them individually as well as compare them to assess efficiency, safety and effectiveness. For their part, innovators who contribute data about their patient flow automation system would receive aggregated insights drawn from the patient flow automation market as a whole. This could help incentivise public and private sector innovators to take part. If that incentive is not enough, cooperation could be also compelled through the inclusion of specific requirements in contracts with patient flow automation providers (for more on this, see the recommendations in use case #1).

This type of data institution would also make it easier for evaluators to perform long-term or periodic evaluations. In this use case, as well as in the others, we identified challenges related to how to perform evaluations not just of a single technology at a single point in time, but how to perform evaluations of many

technologies and/or over long periods of time. Periodic evaluations will be needed for many technologies, in particular technologies that include algorithms that learn ‘on the job’ and therefore modify their decision-making over time. A data institution would provide the stability necessary to enable these types of long-term evaluations.

The data institution would be able to do more than facilitate evaluations, incentivise cooperation and distribute aggregated insights, however. The institution should also be able to help improve the operation of patient flow automation tools and ultimately help the ecosystem function more effectively. By leveraging its position as an intermediary in the ecosystem, the data institution could help to develop and distribute standards – thereby improving the interoperability of the different systems, datasets and technologies, as well as spreading best practice. These standards would need to be developed collaboratively by organisations and stakeholders across the ecosystem, and could aim to define minimum technical, legal and ethical standards that enable interoperability and cross-comparison, while retaining the dynamism that is crucial to efforts to innovate at pace.

Investigate the form and scale of a data institution

We recommend continued research into the form and scale of the proposed data institution, in particular to investigate the following points:

- **To what degree would the data institution need to be independent from the health sector?** It is unclear, for instance, whether the teams developing patient flow automation tools and the hospitals that would be asked to report statistics about those tools would be comfortable reporting to an organisation with close ties to the NHS. During our research, some of our interviewees referred to ‘parent-child syndrome’ and questioned whether organisations, be they public or private sector, would be as willing to report statistics about the operation of their services if there is a perception that the body they are reporting to could use that information to negatively impact them in some way – such as through the loss of contracts or loss of funding. If that is found to be the case, then the independence of the proposed data institution will be of paramount importance. An additional question, then, would be whether the data institution will need to be a newly formed body or if the same roles can be played by an existing, independent organisation.
- **At what scale should the proposed data institution operate and what should be included in its remit?** As with the data institution proposed in use case #2, there are questions related to whether a data institution of this sort should focus on a specific emerging technology, a general type of technology or data technologies more broadly. Similarly, there are questions related to whether such an institution should focus on a specific region or on the UK as a whole. It should be noted that this type of data institution may prove valuable for evaluating many emerging health technologies once they reach a level of adoption where comparative analyses become necessary. Digital-first primary care systems, for instance, are likely to reach this level soon – if they haven’t already. By piloting a data institution focused on patient flow automation tools, it may be possible to design an approach that could be useful for other technologies such as digital-first primary care.

Several people we spoke to suggested that organisations interested in exploring or piloting a data institution along these lines should get in touch with [Clinical Practice](#)

[Research Datalink \(CPRD\)](#) to see if their approach to data collection and stewardship could provide insights or be adapted for the purposes of evaluations. In brief, CPRD collects data from GP practices across the UK and, working with NHS Digital, links that data to a range of other health-related data. They then enable access to that data for various stakeholders (universities, research organisations, charities, pharmaceutical companies), while providing some data and benchmarking insights back to the contributing practices.

Along similar lines, some interviewees suggested that a next step could be to contact the [Research Surveillance Centre](#), which collects data from a range of different platforms and software systems used by GP practices, and therefore has experience in helping to standardise data and data practices.⁴²

Covid-19 context

In the context of Covid-19, having a high-level view of patients and resources within a health system has proved to be crucial.

The government used data from various sources, including confidential UK patient information, in order to obtain this high-level view and as part of their response to the outbreak.⁴³ Rules were adapted to allow for easier sharing of such confidential information.⁴⁴

However, it is unclear whether patients will still be comfortable with data about them being shared so easily in a post-Covid-19 world.

It is also unclear whether this level of sharing should become standard practice in the future. A 'top-down' approach has proved to be valuable in a time of crisis, however going back to a 'bottom-up' approach might be the solution in the long term.

⁴² Royal College of General Practitioners, '[RCGP Research and Surveillance Centre](#)'.

⁴³ The Guardian (2020), '[UK government using confidential patient data in coronavirus response](#)'.

⁴⁴ British Medical Journal (2020), '[Covid-19: Rules on sharing confidential patient information are relaxed in England](#)'.

Conclusions and recommendations

Through our research, we have identified a series of challenges, and the use cases have helped us to dig deep into these challenges. Difficulties in accessing, linking and analysing data from across the health sector, as well as accessing data held by commercial organisations, were highlighted.

Different forms of data institutions are relevant or useful for different types of challenges. In some cases, we have identified a data institution as a way to address current barriers to accessing commercial data, as well as data from across the NHS and the health sector. In other cases, a data institution could help enable comparative and long-term evaluations. We have also explored other areas such as data portability to collect specific types of data.

Working to enact change within one use case can help address similar problems in other use cases. Many of the recommended steps for increasing access to data for the purposes of evaluating the use cases in this report would have value far beyond the use cases themselves.

Below are recommendations for specific groups of stakeholders.

Recommendations for evaluators

- 1. Evaluators should consider themselves data institutions and look for ways to steward data and increase access to data for public benefit.** Evaluators regularly collect and acquire access to a wide range of different datasets (public health, academic research, commercial, contextual) in order to do their evaluations. There is potential for evaluators to adopt the role of a data institution, focusing on bringing together various sources of data, and actively looking to repurpose that data – either for use in further evaluations or to support wider research. This would require effort to clean, describe and make the data they currently hold findable for other researchers who would be interested in accessing that data. Some evaluators may also be able to offer analytical expertise or insight. An important next step along this route would be to engage with patients, data holders and potential reusers of data to discuss their views of such an institution.
- 2. Evaluators should use their position as an intermediary in many interactions to encourage stakeholders across the ecosystem to share data, and enable and support them to improve their capability and trustworthiness.** This should facilitate access to data for evaluation purposes, as well as for research purposes, and enable the deployment and innovation of new tools and services.

3. **Evaluators should act as convenors in the sector to create standards for benchmarking technologies**, in cooperation with innovators, policymakers and funders.

Recommendations for funders

4. **Funders should further explore new models for data access through scoping and pilot studies in the areas outlined above.** Our research has scratched the surface of the challenges of accessing data in these use cases. There are many aspects of new models of data stewardship for these purposes that still need to be explored through testing. The next stages of exploring these models – scoping and piloting – will take time, effort and resources. Seed funding is important to take these efforts to the next stage.⁴⁵
5. **Funders should explore other use cases in the sector such as precision medicine or the impact of social media on mental health.** For precision medicine, it would be interesting to look at the impact of technologies such as AI-based symptom assessment on the health of patients or the work and efficiency of doctors, while also evaluating whether such technologies place additional, unforeseen strain on other parts of the health and care system. Evaluating the impact of social media platforms on the mental health of young people would allow the exploration of the impact that technologies not traditionally understood as ‘health technologies’ have on the health of the population and on the delivery of health and care within the UK.

Recommendations for innovators

6. **Innovators should get evaluators in the room early to arrange the data collection.** Challenges have been identified around a lack of relevant data being collected for the purpose of evaluation. This can be due to evaluation being thought of or put together too late.
7. **Innovators should be prepared to share data for research and evaluation purposes.** There can be some fears related to sharing data as an innovator. However, a key incentive is that sharing of data is necessary for evaluation and research to be successful and insightful, to improve products and to demonstrate their effectiveness for customers.
8. **Innovators should explore best practices around collecting sensitive data about who uses digital services.** Collecting, using and publishing data about who uses digital services is important for demonstrating their effectiveness across different communities and ensuring new technologies do not increase existing inequalities.⁴⁶
9. **Innovators should work together to develop standards for benchmarking and comparison.** This would allow for evaluations to be even more insightful when they can compare different technologies.

⁴⁵ Open Data Institute (2020), ‘[Designing sustainable data institutions](#)’.

⁴⁶ Open Data Institute (2020), ‘[Protected Characteristics in Practice](#)’.

Recommendations for health and care providers

10. **Health and care providers should convene innovators and healthcare practitioners.** This would allow alignment in understanding and expectations related to data collection for evaluation purposes.
11. **Health and care providers should build in evaluation from the start when piloting or deploying new health technologies.** This would allow the sector to keep innovating at pace, while ensuring safety and effectiveness.
12. **Health and care providers should clarify the ways the data will be collected, accessed, used and shared at the procurement stage.** Requirements should be made clear during procurement,⁴⁷ to allow for access to data for evaluation as well as other purposes. In some cases, it may be necessary to embed mandatory access to data for evaluation purposes within procurement contracts.

Recommendations for patient and practitioner groups

13. **Groups representing the users of health technologies (including patients, carers and healthcare practitioners) should explore cooperative models for collecting data about their experience.** Data about stakeholder perspectives is an important part of any evaluation. The formation and stewardship of datasets that capture these experiences and perspectives would allow expectations and concerns to be heard, and for evaluations to be more insightful. Presenting data and facts can be an important way of being heard. These approaches have been explored in other sectors, for example, [Workers Info Exchange](#) helps Uber drivers to collate data about their experience.

⁴⁷ Open Data Institute, '[Guide - How to embed open data into the procurement of public services](#)'.

Appendices

Appendix 1: Project methodology

The ODI started with some desk research on the following topics:

- types of new health technologies
- evaluation approaches for health technology
- data needs and challenges to support evaluation
- health policy data landscape
- data stewardship approaches for health data.

Three workshops were organised to explore each use case. During these workshops, participants were asked to reflect on:

- what they would want to assess or demonstrate through an evaluation, and what data would be necessary to do so
- what the limitations, pain points or challenges of current systems and processes are for gaining access to data, and why they exist.

A wide range of participants took part in these workshops:

- Workshop 1: Babylon Health, AskmyGP, eConsult, Ada Health, UseMyData, NHS England, NHSX
- Workshop 2: Alan Turing Institute, University of Essex, Carnegie Trust, University College of Dublin, King's College London, FullFact, DEMOS/ Centre for the Analysis of Social Media, Department of Health and Social Care, Economic and Social Research Council, Royal Society for Public Health, Nesta, Ada Lovelace Institute, Ipsos Mori
- Workshop 3: Beautiful Information, Future Care Capital, Cardiff University, Google Health, Liverpool Clinical Commissioning Group, Sheffield Teaching Hospitals NHS Foundation Trust, East Midlands Academic Health Science Network.

The ODI also ran some interviews to explore the use cases a bit more and to test some findings. The following organisations took part in the interviews:

- Academic Health Science Networks
- National Data Guardian
- University of Siegen
- Beautiful Information
- Salus Coop
- HealthBank Cooperative
- Alpha Health
- NHSX
- Use My Data
- Ada Lovelace Institute
- Facebook
- Social Science One

- Mumsnet
- Understanding Patient Data
- Health Data Research UK
- Clinical Practice Research Datalink

Appendix 2: Detailed findings

Use case #1: Detailed findings

What an evaluation would aim to assess or demonstrate

Our research found that when evaluating digital-first primary care (DFPC) technologies, evaluators would want to perform a three-pronged evaluation: the measurement of outcomes; the description and analysis of the process; and the sampling of multiple stakeholder perspectives.⁴⁸

Outcomes would cover efficiency of the system as a whole, quality and safety, and effectiveness of care. They could also cover economic aspects (translating impact into economic terms). Analysing the process would mean evaluating how the technology was or is being implemented. Looking at key stakeholders' perspectives would mean looking into patients' experiences as well as staff feedback on the use of this particular technology.⁴⁹

Data an evaluation would need

Evaluators would need access to a wide range of data:

- General practitioner (GP) data
 - Feedback from GP staff on their experience once the technology was implemented
- Patient data
 - Data on patients' backgrounds in relation to their access and use of technology in general (for example, deprivation, digital literacy, internet connection)
 - Patient demographic data (for example, age, gender, ethnicity)
 - Usage data (data related to how the patients used the technology)
 - Patient pathways (unnecessary accident and emergency attendance)
 - Patient outcomes (what happened after the online consultation? Did the patient receive health advice? Was it followed by a face-to-face or telephone consultation?)
 - Feedback on patients' experience using the technology (user satisfaction and understanding)
- Operational data
 - Aggregated utilisation data related to how and when the digital access was used by patients (for example, aggregate data from all uses of the technology in order to understand how representative they were)

⁴⁸ Ellis R, Hogard E (2006), '[The Trident: A Three-Pronged Method for Evaluating Clinical, Social and Educational Innovations](#)'.

⁴⁹ The Health Foundation and the Open Data Institute, '[Workshop 1 notes](#)'.

- Number of online consultations and triage data
- NHS system-level data (for example, waiting times, GP workflow)
- Data about the decision-making process used for the technology (algorithms as well as any data used to train them).

There is also a need to link these datasets together, to understand the pathways taken by patients through the system, and the wider impact of the service, for example:

- Data between the provider, GP practices and other health services (eg Secondary Use Service⁵⁰) to understand outcomes and show how they might have changed.
- There could also be an ability to link familial records if patients are happy to consent to it.

Challenges or barriers to accessing necessary data

An evaluation would require the gathering of data from multiple sources. It can prove to be difficult to access this data, for the following reasons.

Some data is not currently being collected:

- There may be a lack of data, limiting the effectiveness of evaluation of DFPC technology. This can be related to a lack of resources (time and capacity), high data collection costs or even a particular data collection not being set as a priority (for example, surveys, or user or GP experience statistics).
- The collection and/or analysis is often not financially feasible.

Accessing the data can be difficult:

- There may be concerns related to data sharing, depending on the data steward, such as:
 - Tech providers may have legal and intellectual property concerns.
 - GPs may be concerned about who is accessing their patient data and for what purposes.
 - Local authorities may have concerns about privacy and may face technical barriers to sharing.
 - Individuals may have concerns about their own privacy.
- Data may not be shared or accessible to a certain extent because:
 - there may be a lack of real-time data
 - the data may not be granular enough
 - the data may be anonymised to the point that it is less useful
 - the data may only be shared for a brief period of time, however evaluation requires regular or continued access.

When accessing the data, it can lack consistency:

- There may be gaps in the datasets (collected with a different aim in mind). There is no single, national primary care dataset that can be used for research or evaluation.
- The data may not be of a high enough quality (and therefore difficult to combine or interoperate).

⁵⁰ NHS Digital, '[Secondary Uses Service \(SUS\)](#)'.

- Data may not linkable (for example across a disease rather than across an individual; between DFPC provider and GP clinical system; across primary and secondary care)
 - DFPC providers are unable to link their data with national datasets such as Secondary Uses Service/Hospital Episode Statistics.⁵¹
 - Different GP practices use different patient record systems. Not all DFPC systems necessarily link in to GP-held patient records.

Use case #2: Detailed findings

What an evaluation would aim to assess or demonstrate

Understanding correlation and/or causation:

- What information is present online?
- What information have people been exposed to?
- What external factors influence people's interactions with such information (eg socioeconomic background, the source)?
- What is the correlation between what people see online and their beliefs/behaviour?

Understanding the people actively playing a role in spreading misinformation and the ones being impacted by it:

- Understanding the intent of people propagating misinformation about vaccines.
- Understanding the communities within which the information is flowing:
 - What makes people more susceptible to misinformation?
 - What is their exposure (information related to their identity, social organisation, the power relationships they are involved in)?

Measuring and understanding the impact misinformation has on people and how to challenge misinformation:

- Identifying its direct impact on vaccination rates. This could help identify what the highest impact interventions could be (for instance, how effective are counter messages, do they make a difference to behaviours?).
- Understanding what steps have been proven capable of combating either the spread of misinformation or user beliefs/behaviour.
- Understanding where trust exists (for instance with authority figures such as local GPs, central authorities): what factors impact the level of trust that people place in information and the source of that information (for example, when the information is coming from the NHS, versus a politician, versus a friend)?

Evaluators may also want to have a view on what platforms and other sources of online information are doing that impacts the spread of misinformation and user behaviour:

- What, in the design and algorithms being used for instance, makes the spread of misinformation easier?

⁵¹ NHS Digital, '[Hospital Episode Statistics](#)'.

- When misinformation is identified, what actions are put in place to fight against it?

Data an evaluation would need

To assess some of the elements listed above, various types of data would be needed:

- Data about the content
 - Data related to misinformation being available online: measuring what content actually exists on platforms, how much and where.
- Data about engagement with misinformation
 - Data about how people engage with misinformation: metadata about how it is viewed, shared and commented on.
 - Data on engagement with offline sources of official information (eg GP and health visitor interactions, public health communications), as well as data related to how misinformation online interacts with misinformation offline.
- Data about people
 - Data about people's background: information about identities (belonging, social organisation underlying beliefs).
 - Data on the impact and profile of the personal relationships that have the most influence: "friends of friends who think x", "mums networks".
 - Data about the impact on behaviours and health outcomes: is misinformation leading to a change in attitude? Is it leading to a change in vaccine uptake?
- Granular data related to geographies (to allow for an analysis at city level for instance).

Challenges or barriers to accessing necessary data

There are various challenges related to accessing the data that would help assess the impact of online misinformation on vaccine hesitancy.

Some data is currently not being collected on this topic:

- There can be various reasons for some data not being collected at the moment.
 - Organisations may:
 - just not want to collect data
 - lack time and/or resources
 - lack infrastructure
 - be barred from collecting it
 - not realise it would be useful
 - be uncertain about whether they can collect it
 - People may not be willing to have some data collected.
- There is currently limited follow up with people who do not present for routine vaccinations.

- There is a lack of insight and knowledge (from government, NHS, public health organisations) on patients' behaviours and motivations around vaccination.

The topic is complex and so the data needed to analyse it is complex:

- There are challenges related to knowing what data exists, who holds it and how to access it.
- The activity on an online account does not capture the experience of a person. The context both across the online space and offline is important.
- Data on vaccination behaviours has limitations. Understanding all the factors leading to vaccine hesitancy can be challenging.

Accessing the data can be difficult:

- Data on online behaviours and the algorithms used is held by private tech companies, who may not always be willing to share the data.
 - This may be due to:
 - commercial sensitivity and intellectual property (some data may contain strategic information, such as advertising efficacy, or data about the algorithms used for instance)
 - privacy concerns
 - ethics concerns related to the use of personal data (it may be that the platforms are very wary of sharing any information of a personal nature and users are wary of them doing so).
 - It might also be due to a lack of incentives: why would a platform give such deep access if it is not required? What is the benefit for them? This is especially the case if the data requested is core to their business, intellectual property or unique selling proposition.
- There is a lack of clarity around what is and is not permissible. Some stakeholders may need clarity on what can be done and what data can be accessed. Otherwise, data may not be shared when it could legally and ethically be shared, and data stewards may be uncertain and therefore wary of running afoul of laws/regulations that are not clear.
- Data can also be held by individuals, and they may have privacy concerns related to the use of data about them.
- Accessing the data may be costly.

When accessing the data, there may be a lack of consistency:

- There is a lack of consistency or standardisation across platforms for flagging, labelling, reporting and removing misinformation, which will complicate any analysis across different online sources (and such processes might change rapidly within single platforms). This is because most misinformation is likely to be contested or have subjective aspects.
- The format in which data is accessible might not always be useful.
- Linking data between online platforms and health services is difficult.

Use case #3: Detailed findings

What an evaluation would aim to assess or demonstrate

- Impact on patients:

- What is the patient experience?
- What are the patient outcome consequences (for example, cancer pathways are normally set within timeframes and relate to patient flow across a health system from primary to secondary care)?
- Are patients accessing other services (for example, are mental health patients also accessing community services?) and what is the broader health profile of patients accessing these services?
- Are there any unintended consequences (for example, fairness)?
- Impact on the whole system:
 - How does it impact social care, GPs and community care (for example, are decisions made by an AI-backed patient pathway tool causing added strain on a particular section of the health and care system)?
 - How does it impact the quality of care and where are the anomalies (for example, where are the 3% of patients suffering an extraordinary wait)?

Data an evaluation would need

To assess or demonstrate all of the above, a wide range of data would be needed. It would come from various services such as hospitals, mental health services, community GPs, as well as providers such as [EMIS](#), and national services such as the Office for National Statistics (ONS; for census data).

- Patient data
 - Measures of patient satisfaction.
- Operational data
 - Data from across the system (for example, were patients shunted elsewhere? Would the receiving service have capacity?).
- Provider data
 - Data on use of the tool
 - Training data (eg data about the original context), particularly if an algorithm is involved and has been trained in a specific context for a specific purpose.

Some identifiable data would be needed, especially to be able to link data together. Data would also need to be comparable across hospitals/trusts.

Challenges or barriers to accessing necessary data

There are various challenges related to accessing the data that would help assess the impact of patient flow tools.

Accessing the data can be difficult:

- Obtaining data from various companies can be an issue. This can be due to:
 - a lack of time and resources
 - a lack of incentives
 - a cultural reluctance to share data (within care providers and across them)

- Patient safety teams would want their data to be understood in the right way. Some might have worries around the use of their data.
 - There can be patient and public reluctance as well.
 - There can be trust issues between patients and the public and private sectors.
- Data may not be shared or accessible to a certain extent
 - Time lags can be a challenge. This can be due to coding and data quality processes needed before the data can be used.

When accessing the data, it can lack consistency:

- There are various clinical systems and trusts, where data collection is inconsistent and not standardised.
- There are data quality risks related to manual input (which could be related to resources and technical issues), for example, patient data being manually entered on spreadsheets is prone to lag, duplication and error. This is the case for care homes for instance, where community care data is often recorded manually and not shared.
 - Related to this, there is a lack of analytical expertise to represent data in such a way that facilitates improvement measures.
- There is no defined terminology or taxonomy for data collectors to describe the data to consumers.
- Data without context can be problematic: without understanding the original purpose, interpreting it for secondary use can be challenging.

There can also be challenges related to people not knowing what is permissible (for example, people not being comfortable with linking data, as they are not confident they are doing it right). It is also sometimes difficult to know who controls different datasets – for instance, is it the hospital that collected the data or the software provider that built the data-collection system?